STUDIES IN STATISTICAL INFERENCE, SAMPLING TECHNIQUES AND DEMOGRAPHY

Rajesh Singh

Department of Statistics, BHU, Varanasi (U. P.), India

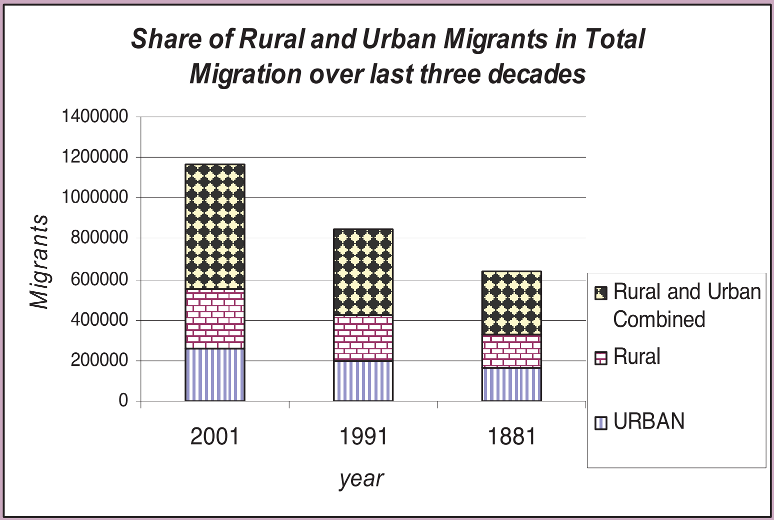
Jayant Singh

Department of Statistics

Rajasthan University, Jaipur, India

Florentin Smarandache

Chair of Math & Sciences Dept., University of New Mexico, Gallup, USA



STUDIES IN STATISTICAL INFERENCE, SAMPLING TECHNIQUES AND DEMOGRAPHY

Rajesh Singh, Jayant Singh, Florentin Smarandache

ILQ

2009

This book can be ordered in a paper bound reprint from:

Books on Demand

ProQuest Information & Learning

(University of Microfilm International)

300 N. Zeeb Road

P.O. Box 1346, Ann Arbor

MI 48106-1346, USA

Many books can be downloaded from the following

Digital Library of Science:

http://www.gallup.unm.edu/~smarandache/eBooks-otherformats.htm

Peer Reviewers:

Prof. Mihaly Bencze, Department of Mathematics, Áprily Lajos College, Braşov, Romania.

Dr. Sukanto Bhattacharya, Department of Business Administration, Alaska Pacific University, U.S.A.

(ISBN-10): 1-59973-087-1

(ISBN-13): 978-1-59973-087-5

(EAN): 9781599730875

Printed in the United States of America

Contents

Preface

This volume is a collection of five papers. Two chapters deal with problems in statistical inference, two with inferences in finite population, and one deals with demographic problem. The ideas included here will be useful for researchers doing works in these fields. The following problems have been discussed in the book:

Chapter 1. In this chapter optimum statistical test procedure is discussed. The test procedures are optimum in the sense that they minimize the sum of the two error probabilities as compared to any other test. Several examples are included to illustrate the theory.

Chapter 2. In testing of hypothesis situation if the null hypothesis is rejected will it automatically imply alternative hypothesis will be accepted? This problem has been discussed by taking examples from normal distribution.

Chapter 3. In this section improved chain-ratio type estimator for estimating population mean using some known values of population parameter(s) has been discussed. The proposed estimators have been compared with two-phase ratio estimator and some other chain ratio type estimators.

Chapter 4. In this section we have analysed exponential ratio and exponential product type estimators using two auxiliary variables are proposed for estimating unknown population variance S_y^2 . Problem is extended to the case of two-phase sampling.

Chapter 5. In this section structural dynamics of various causes of migration in Jaipur was analysed. Reasons of migration from rural to urban areas and that of males and females are studied.

The Authors

Optimum Statistical Test Procedure

Rajesh Singh

Department of Statistics, BHU, Varanasi (U.P.), India

rsinghstat@yahoo.com

Jayant Singh

Department of Statistics

Rajasthan University, Jaipur, India

Jayantsingh47@rediffmail.com

Florentin Smarandache

Chair of Department of Mathematics, University of New Mexico, Gallup, USA

smarand@unm.edu

Introduction

Let X be a random variable having probability distribution $P(X/\theta)$, $\theta \in \Theta$. It is desired to test $H_0 : \theta \in \Theta_0$ against $H_1 : \theta \in \Theta_1 = \Theta - \Theta_0$. Let S denote the sample space of outcomes of an experiment and $\underline{x} = (x_1, x_2, ---, x_n)$ denote an arbitrary element of S. A test procedure consists in diving the sample space into two regions W and $S - W$ and deciding to reject H_0 if the observed $\underline{x} \in W$. The region W is called the critical region. The function $\gamma(\theta) = P_\theta(\underline{x} \in W) = P_\theta(W)$, say, is called the power function of the test.

We consider first the case where Θ_0 consists of a single element, θ_0 and its complement Θ_1 also has a single element θ_1. We want to test the simple hypothesis $H_0 : \theta = \theta_0$ against the simple alternative hypothesis $H_1 : \theta = \theta_1$.

Let $L_0 = L(X/H_0)$ and $L_1 = L(X/H_1)$ be the likelihood functions under H_0 and H_1 respectively. In the Neyman – Pearson set up the problem is to determine a critical region W such that

$$\gamma(\theta_0) = P_{\theta_0}(W) = \int_W L_0 dx = \alpha, \text{ an assigned value} \tag{1}$$

and $$\gamma(\theta_1) = P_{\theta_1}(W) = \int_W L_1 dx \quad \text{is maximum} \tag{2}$$

compared to all other critical regions satisfying (1).

If such a critical region exists it is called the most powerful critical region of level α.

By the Neyman-Pearson lemma the most powerful critical region W_0 for testing $H_0: \theta = \theta_0$ against $H_1: \theta = \theta_1$ is given by

$$W_0 = \{\underline{x} : L_1 \geq kL_0\}$$

provided there exists a k such that (1) is satisfied.

For this test $\gamma(\theta_0) = \alpha$ and $\gamma(\theta_1) \to 1$ as $n \to \alpha$.

But for any good test we must have $\gamma(\theta_0) \to 0$ and $\gamma(\theta_1) \to 1$ as $n \to \infty$ because complete discrimination between the hypotheses H_0 and H_1 should be possible as the sample size becomes indefinitely large.

Thus for a good test it is required that the two error probabilities α and β should depend on the sample size n and both should tend to zero as $n \to \infty$.

We describe below test procedures which are optimum in the sense that they minimize the sum of the two error probabilities as compared to any other test. Note that minimizing $(\alpha + \beta)$ is equivalent to maximising

$$1 - (\alpha + \beta) = (1 - \beta) - \alpha = \text{Power} - \text{Size}.$$

Thus an optimum test maximises the difference of power and size as compared to any other test.

Definition 1 : A critical region W_0 will be called optimum if

$$\int_{W_0} L_1 dx - \int_{W_0} L_0 dx \geq \int_{W} L_1 dx - \int_{W} L_0 dx \tag{3}$$

for every other critical region W.

Lemma 1: For testing $H_0: \theta = \theta_0$ against $H_1 : \theta = \theta_1$ the region

$W_0 = \{\underline{x} : L_1 \geq L_0\}$ is optimum.

Proof : W_0 is such that inside W_0, $L_1 \geq L_0$ and outside W_0, $L_1 < L_0$. Let W be any other critical region.

$$\int_{W_0} (L_1 - L_0) dx - \int_{W} (L_1 - L_0) dx$$

$$= \int_{W_0 \cap W^c} (L_1 - L_0) dx - \int_{W \cap W_0^c} (L_1 - L_0) dx,$$

by subtracting the integrals over the common region $W_0 \cap W$.

≥ 0

since the integrand of first integral is positive and the integrand under second integral is negative.

Hence (3) is satisfied and W_0 is an optimum critical region.

Example 1 :

Consider a normal population $N(\theta, \sigma^2)$ where σ^2 is known.

It is desired to test H_0: $\theta = \theta_0$ against H_1 : $\theta = \theta_1$, $\theta_1 > \theta_0$.

$$L(x/\theta) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n e^{-\sum_{i=1}^{n}\frac{(x_i-\theta)^2}{2\sigma^2}}$$

$$\frac{L_1}{L_0} = \frac{e^{-\frac{\sum(x_i-\theta_1)^2}{2\sigma^2}}}{e^{-\frac{\sum(x_i-\theta_0)^2}{2\sigma^2}}}$$

The optimum test rejects H_o

if $\quad \frac{L_1}{L_0} \geq 1$

i.e. if $\quad \log\frac{L_1}{L_0} \geq 0$

i.e. if $\quad -\frac{\sum(x_i-\theta_1)^2}{2\sigma^2} + \frac{\sum(x_i-\theta_0)^2}{2\sigma^2} \geq 0$

i.e. if $\quad 2\theta_1\sum x_i - n\theta_1^2 - 2\mu_0\sum x_i + n\theta_0^2 \geq 0$

i.e. if $\left(2\theta_1 - 2\theta_0\right)\sum x_i \geq n\left(\theta_1^2 - \theta_0^2\right)$

i.e. if $\dfrac{\sum x_i}{n} \geq \dfrac{\theta_1 + \theta_0}{2}$

i.e. if $\bar{x} \geq \dfrac{\theta_1 + \theta_0}{2}$

Thus the optimum test rejects H_0

if $\bar{x} \geq \dfrac{\theta_1 + \theta_0}{2}$

We have

$$\alpha = P_{H_0}\left[\bar{x} \geq \frac{\theta_1 + \theta_0}{2}\right]$$

$$= P_{H_0}\left[\frac{\bar{x} - \theta_0}{\sigma / \sqrt{n}} \geq \frac{\sqrt{n}\left(\theta_1 - \theta_0\right)}{2\sigma}\right]$$

Under H_o, $\dfrac{\bar{x} - \theta_0}{\left(\sigma / \sqrt{n}\right)}$ follows N(0,1) distribution.

$$\therefore \quad \alpha = 1 - \Phi\left(\frac{\sqrt{n}\left(\theta_1 - \theta_0\right)}{2\sigma}\right)$$

where Φ is the c.d.f. of a N(0,1) distribution.

$$\beta = P_{H_1}\left[\bar{x} < \frac{\theta_1 + \theta_0}{2}\right] = P_{H_1}\left[\frac{\bar{x} - \theta_1}{\sigma / \sqrt{n}} < \frac{-\sqrt{n}\left(\theta_1 - \theta_0\right)}{2\sigma}\right]$$

Under H_1, $\dfrac{\bar{x}-\theta_1}{\left(\sigma/\sqrt{n}\right)}$ follows N(0,1) distribution.

$$\beta = 1-\Phi\left(\frac{\sqrt{n}(\theta_1-\theta_0)}{2\sigma}\right)$$

Here $\alpha = \beta$.

It can be seen that $\alpha = \beta \to 0$ as $n \to \infty$.

Example 2 : For testing H_0: $\theta = \theta_0$ against H_1 : $\theta = \theta_1$, $\theta_1 < \theta_0$, the optimum test rejects

H_0 when $\bar{x} \le \dfrac{\theta_1+\theta_0}{2}$.

Example 3 : Consider a normal distribution $N(\theta, \sigma^2)$, θ known.

It is desired to test H_0 : $\sigma^2 = \sigma_0^2$ against H1 : $\sigma^2 = \sigma_1^2, \sigma_1^2 > \sigma_0^2$.

We have

$$L(x/\sigma^2) = \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{n}{2}} e^{-\sum\frac{(x_i-\theta)^2}{2\sigma^2}}$$

$$\frac{L_1}{L_0} = \frac{\left(\dfrac{\sigma_0^2}{\sigma_1^2}\right)^{\frac{n}{2}} e^{-\frac{\sum(x_i-\theta)^2}{2\sigma_1^2}}}{e^{-\frac{\sum(x_i-\theta)^2}{2\sigma_0^2}}}$$

$$\log\frac{L_1}{L_0} = -\frac{n}{2}\left(\log\sigma_1^2 - \log\sigma_0^2\right) - \frac{\sum(x_i-\theta)^2}{2\sigma_1^2} + \frac{\sum(x_i-\theta)^2}{2\sigma_0^2}$$

$$= -\frac{n}{2}\left(\log\sigma_1^2 - \log\sigma_0^2\right) + \frac{\sum(x_i-\theta)^2}{2}\frac{\sigma_1^2-\sigma_0^2}{\sigma_1^2\sigma_0^2}$$

The optimum test rejects H_0

$$if \quad \frac{L_1}{L_0} \geq 1$$

i.e. if $\displaystyle \frac{\sigma_1^2-\sigma_0^2}{2\sigma_1^2\sigma_0^2}\sum(x_i-\theta)^2 \geq \frac{n}{2}\left(\log\sigma_1^2 - \log\sigma_0^2\right)$

i.e. if $\displaystyle \frac{\sum(x_i-\theta)^2}{\sigma_0^2} \geq \frac{n\sigma_1^2}{\sigma_1^2-\sigma_0^2}\left(\log\sigma_1^2 - \log\sigma_0^2\right)$

i.e. if $\displaystyle \sum\left(\frac{x_i-\theta}{\sigma_0}\right)^2 \geq nc$

where $\displaystyle c = \frac{\sigma_1^2}{\sigma_1^2-\sigma_0^2}\left(\log\sigma_1^2 - \log\sigma_0^2\right)$

Thus the optimum test rejects H_0 if $\sum\left(\frac{x_i-\theta}{\sigma_0}\right)^2 \geq \; nc$.

Note that under $H_0 : \dfrac{x_i-\theta}{\sigma_0}$ follows N(0,1) distribution. Hence $\displaystyle\sum\left(\frac{x_i-\theta}{\sigma_0}\right)^2$ follows, under H_0, a chi-square distribution with n degrees of freedom (d. f.).

Here $\displaystyle \alpha = P_{H_0}\left[\sum\left(\frac{x_i-\theta_0}{\sigma_0}\right)^2 \geq nc\right] = P\left[\chi^2_{(n)} \geq nc\right]$

and $\displaystyle 1-\beta = P_{H_1}\left[\sum\left(\frac{x_i-\theta}{\sigma_0}\right)^2 \geq nc\right]$

$$= P_{H_1}\left[\sum\left(\frac{x_i-\theta}{\sigma_1}\right)^2 \geq \frac{nc\sigma_0^2}{\sigma_1^2}\right]$$

$$= P_{H_1}\left[\chi^2_{(n)} \geq \frac{nc\sigma_0^2}{\sigma_1^2}\right]$$

Note that under H1, $\sum\left(\frac{x_i-\theta}{\sigma_1}\right)^2$ follows a chi-square distribution with n d.f.

It can be seen that $\alpha \to 0$ and $\beta \to 0$ as $n \to \infty$.

Example 4 : Let X follow the exponential family distributions

$$f(x/\theta) = c(\theta)e^{Q(\theta)T(x)}h(x)$$

It is desired to test H_0: $\theta = \theta_0$ against H_1 : $\theta = \theta_1$

$$L(x/\theta) = [c(\theta)]^n e^{Q(\theta)\sum T(x_i)}\prod_i h(x_i)$$

The optimum test rejects H_o when

$$log\frac{L_1}{L_0} \geq 0$$

i.e. when $[Q(\theta_1) - Q(\theta_0)]\sum T(x_i) \geq n\log\frac{c(\theta_0)}{c(\theta_1)}$

i.e. when $\sum T(x_i) \geq \dfrac{n\log\frac{c(\theta_0)}{c(\theta_1)}}{[Q(\theta_1) - Q(\theta_0)]}$ if $Q(\theta_1) - Q(\theta_0) > 0$

and rejects H_o,

$$\text{when} \quad \sum T(x_i) \le \frac{n \log \dfrac{c(\theta_0)}{c(\theta_1)}}{[Q(\theta_1) - Q(\theta_0)]} \quad \text{if } Q(\theta_1) - Q(\theta_0) < 0$$

Locally Optimum Tests:

Let the random variable X have probability distribution $P(x/\theta)$. We are interested in testing $H_0: \theta = \theta_0$ against $H_1: \theta > \theta_0$. If W is any critical region then the power of the test as a function of θ is

$$\gamma(\theta) = P_\theta(W) = \int_W L(X/\theta)dx$$

We want to determine a region W for which

$$\gamma(\theta) - \gamma(\theta_0) = \int_W L(x/\theta)dx - \int_W L(x/\theta)$$

is a maximum.

When a uniformly optimum region does not exist, there is not a single region which is best for all alternatives. We may, however, find regions which are best for alternatives close to the null hypothesis and hope that such regions will also do well for distant alternatives. We shall call such regions locally optimum regions.

Let $\gamma(\theta)$ admit Taylor expansion about the point $\theta = \theta_0$. Then

$$\gamma(\theta) = \gamma(\theta_0) + (\theta - \theta_0)\acute{\gamma}(\theta_0) + \delta \qquad \text{where } \delta \to 0 \text{ as } \theta \to \theta_0$$

$$\therefore \quad \gamma(\theta) - \gamma(\theta_0) = (\theta - \theta_0)\acute{\gamma}(\theta_0) + \delta$$

If $|\theta - \theta_0|$ is small $\gamma(\theta) - \gamma(\theta_0)$ is maximised when $\acute{\gamma}(\theta_0)$is maximised.

Definition2 : A region W_0 will be called a locally optimum critical region if

$$\int_{W_0} \acute{L}(x/\theta_0)dx \geq \int_W \acute{L}(x/\theta_0)dx \quad (4)$$

For every other critical region W.

Lemma 2 : Let W_0 be the region $\{\underline{x}: \acute{L}(x/\theta_0) \geq L(x/\theta_0)\}$. Then W_0 is locally optimum.

Proof: Let W_0 be the region such that inside it $\acute{L}(x/\theta_0) \geq L(x/\theta_0)$ and outside it $\acute{L}(x/\theta_0) < L(x/\theta_0)$. Let W be any other region.

$$\int_{W_0} \acute{L}(x/\theta_0)dx - \int_W \acute{L}(x/\theta_0)dx$$

$$= \int_{W_{0\cap W^c}} \acute{L}(x/\theta_0)dx - \int_{W_0^c \cap W} \acute{L}(x/\theta_0)dx$$

$$= \int_{W_{0\cap W^c}} \acute{L}(x/\theta_0)dx + \int_{W_0^c \cup W} \acute{L}(x/\theta_0)dx \quad (*)$$

$$\geq \int_{W_{0\cap W^c}} L(x/\theta_0)dx + \int_{W_0^c \cup W} L(x/\theta_0)dx$$

since $\acute{L}(x/\theta_0) \geq L(x/\theta_0)$ inside both the regions of the integrals.

≥ 0, since $L(x/\theta_0) \geq 0$ in all the regions.

Hence $\int_{W_0} \acute{L}(x/\theta_0)dx \geq \int_W \acute{L}(x/\theta_0)dx$ for every other region W.

To prove (*):

We have $\int_R L(x/\theta_0)dx + \int_{R^c} L(x/\theta_0)dx = 1$ for every region R.

Differentiating we have

$$\int_R \acute{L}(x/\theta_0)dx + \int_{R^c} \acute{L}(x/\theta_0)dx = 0$$

$$\int_R \acute{L}(x/\theta_0)dx = -\int_{R^c} \acute{L}(x/\theta_0)dx = 0$$

In (*), take R^C=$W_0^c \cap W$ and the relation is proved.

Similarly if the alternatives are $H_1 : \theta < \theta_0$, the locally optimum critical region is

$$\{\underline{x}: \acute{L}(x/\theta_0) \leq L(x/\theta_0)\}.$$

Example 5: Consider $N(\theta, \sigma^2)$ distribution, σ^2 known.

It is desired to test H_0: $\theta = \theta_0$ against $H_1 : \theta > \theta_0$

$$L(x/\theta) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n e^{\frac{-\sum(x_i-\theta)^2}{2\sigma^2}}$$

$$logL(x/\theta) = nlog\left(\frac{1}{\sigma\sqrt{2\pi}}\right) - \frac{\sum(x_i-\theta)^2}{2\sigma^2}$$

$$\frac{\acute{L}(x/\theta)}{L(x/\theta)} = \frac{\delta logL(x/\theta)}{\delta\theta} = \frac{\sum(x_i-\theta)}{\sigma^2} = \frac{n(\bar{x}-\theta)}{\sigma^2}$$

$$\therefore \quad \frac{\acute{L}(x/\theta_0)}{L(x/\theta_0)} = \frac{n(\bar{x}-\theta_0)}{\sigma^2}$$

The locally optimum test rejects H_0 ,if

$$\frac{n(\bar{x}-\theta_0)}{\sigma^2} \geq 1$$

i.e. $\bar{x} \geq \theta_0 + \frac{\sigma^2}{n}$

Now,

$$\alpha \quad = P_{H_0}\left[\bar{x} \geq \theta_0 + \frac{\sigma^2}{n}\right]$$

$$= P_{H_0}\left[\frac{\bar{x}-\theta_0}{\sigma/\sqrt{n}} \geq \sigma/\sqrt{n}\right]$$

$= 1 - \Phi\left(\sigma/\sqrt{n}\right)$, since under H$_0$, $\frac{\bar{x}-\theta_0}{\sigma/\sqrt{n}}$ follows N(0,1) distribution.

$$1-\beta = P_{H_1}\left[\bar{x} \geq \theta_0 + \sigma^2/n\right]$$

$$= P_{H_1}\left[\frac{\bar{x}-\theta_1}{\sigma/\sqrt{n}} \geq -\frac{\theta_1-\theta_0}{\sigma/\sqrt{n}} + \frac{\sigma}{\sqrt{n}}\right]$$

$$= 1 - \Phi\left[\frac{-(\theta_1-\theta_0\sqrt{n})}{\sigma} + \frac{\sigma}{\sqrt{n}}\right]$$

since under H$_1$, $\frac{\bar{x}-\theta_1}{\sigma/\sqrt{n}}$ follows N(0,1) distribution.

Exercise : If $\theta_0 = 10, \theta_1 = 11, \sigma = 2, n = 16$, then α = 0.3085, 1-β = 0.9337

Power - Size = 0.6252

If we reject H$_0$ when $\frac{\bar{x}-\theta_0}{\sigma/\sqrt{n}}$ > 1.64, then α = 0.05, 1-β = 0.6406

Power – Size = 0.5906

Hence Power – Size of locally optimum test is greater than Power – size of the usual test.

Locally Optimum Unbiased Test:

Let the random variable X follows the probability distribution $P(x/\theta)$. Suppose it is desired to test H_0: $\theta = \theta_0$ against H_1 : $\theta \neq \theta_0$. We impose the unbiasedness restriction

$\gamma(\theta) \geq \gamma(\theta_0), \theta \neq \theta_0$

and $\gamma(\theta) - \gamma(\theta_0)$ is a maximum as compared to all other regions. If such a region does not exist we impose the unbiasedness restriction $\acute{\gamma}(\theta_0) = 0$.

Let $\gamma(\theta)$ admit Taylor expansion about the point $\theta = \theta_0$.

Then $\gamma(\theta) = \gamma(\theta_0) + (\theta - \theta_0)\acute{\gamma}(\theta_0) + \frac{(\theta-\theta_0)^2}{2}\gamma''(\theta_0) + \eta$

where $\eta \to 0$ as $\theta \to 0$.

$\therefore \quad \gamma(\theta) - \gamma(\theta_0) = (\theta - \theta_0)\acute{\gamma}(\theta_0) + \frac{(\theta-\theta_0)^2}{2}\gamma''(\theta_0) + \eta$

Under the unbiasedness restriction $\acute{\gamma}(\theta_0) = 0$, if $|\theta - \theta_0|$ is small $\gamma(\theta) - \gamma(\theta_0)$ is maximised when $\gamma''(\theta_0)$ is maximised.

Definition 3 : A region W_0 will be called a locally optimum unbiased region if

$$\acute{\gamma}(\theta_0) = \int_{W_0} \acute{L}(x/\theta_0)dx = 0 \tag{5}$$

and
$$\gamma''(\theta_0) = \int_{W_0} L''(X/\theta_0)dx \geq \int_{W} L''(X/\theta_0)dx \tag{6}$$

for all other regions W satisfying (5).

Lemma 3 : Let W_0 be the region $\{\underline{x} : L''(x/\theta_0) \geq L(x/\theta_0)\}$

Then W_0 is locally optimum unbiased.

Proof : Let W be any other region

$\int_{W_0} L''(x/\theta_0)dx - \int_W L''(x/\theta_0)dx$

$$= \int_{W_0 \cap W^c} L''(x/\theta_0)dx + \int_{W_0^c \cap W} L''(x/\theta_0)dx$$

by subtracting the common area of W and W_0.

$$= \int_{W_0 \cap W^c} L''(x/\theta_0)dx + \int_{W_0 \ \cap W^c} L''(x/\theta_0)dx,$$

since $L''(x/\theta_0) \geq L(x/\theta)$ inside W_0 and outside W.

$\geq \; 0$ since $L(x/\theta) \geq 0.$

Example 6:

Consider $N(\theta, \sigma^2)$ distribution, σ^2 known .

$H_0{:}\, \theta = \theta_0, \;\; H_1{:}\, \theta \neq \theta_0$

$$L = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n e^{-\frac{\sum(x_i-\theta)^2}{2\sigma^2}}$$

$$\frac{L''(x/\theta)}{L(x/\theta)} = \frac{n^2(\bar{x}-\theta)^2}{\sigma^4} - \frac{n}{\sigma^2}$$

Locally optimum unbiased test rejects H_0

$$if \;\; \frac{n^2(\bar{x}-\theta_0)^2}{\sigma^4} - \frac{n}{\sigma^2} \geq 1$$

$$i.e. \;\; \frac{n(\bar{x}-\theta_0)^2}{\sigma^2} \geq 1 + \frac{\sigma^2}{n}$$

$Under\ H_{0,}$ $\frac{n(\bar{x}-\theta_0)^2}{\sigma^2}$ follows $\chi^2_{(1)}$ distribution.

Testing Mean of a normal population when variance is unknown.

Consider $N(\theta, \sigma^2)$ distribution, σ^2 known.

For testing H_0: $\theta = \theta_0$ against H_1 : $\theta = \theta_1$, the critical function of the optimum test is given by

$$\phi_m(x) = \begin{cases} 1 \ \ if \ L(x/\theta_1) \geq L(x/\theta_0) \\ 0 \ otherwise \end{cases}$$

On simplification we get

$$\phi_m(x) = \begin{cases} 1 \ \ if \ \ \bar{x} \geq \dfrac{\theta_0 + \theta_1}{2} \ if \ \mu_1 > \mu_0 \\ 0 \quad otherwise \end{cases}$$

$$\phi_m(x) = \begin{cases} 1 \ \ \text{if} \ \bar{x} < \dfrac{\theta_0 + \theta_1}{2} \ \text{if} \ \mu_1 < \mu_0 \\ 0 \quad \text{otherwise} \end{cases}$$

Consider the case when σ^2 is unknown.

For this case we propose a test which rejects H_0 when

$$\frac{\hat{L}(x/\theta_1)}{\hat{L}(x/\theta_0)} \geq 1$$

where $\hat{L}(x/\theta_i)$, (i=0,1) is the maximum of the likelihood under H_i obtained from $L(x/\theta_i)$ by replacing σ^2 by its maximum likelihood estimate

$$\widehat{\sigma_i^2} = \frac{1}{n}\sum_{j=1}^{n}\left(x_j - \theta_i\right)^2; \ \text{i=0,1.}$$

Let $\phi_p(x)$ denote the critical function of the proposed test, then

$$\phi_p(x) = \begin{cases} 1 \ \ if \ \hat{L}(x/\theta_1) \ \geq \hat{L}(x/\theta_0) \\ 0 \ \ otherwise \end{cases}$$

On simplification we get

$$\phi_p(x) = \begin{cases} 1 \ \ if \ \bar{x} \geq \frac{\theta_0+\theta_1}{2} & if \ \theta_1 > \theta_0 \\ o \ \ otherwsie & \end{cases}$$

and $$\phi_p(x) = \begin{cases} 1 \ \ if \ \bar{x} < \frac{\theta_0+\theta_1}{2} & if \ \theta_1 < \theta_0 \\ o \ \ otherwsie & \end{cases}$$

Thus the proposed test $\phi_p(x)$ is equivalent to $\phi_m(x)$ which is the optimum test that minimizes the sum of two error probabilities $(\alpha + \beta)$. Thus we see that one gets the same test which minimises the sum of the two error probabilities irrespective of whether σ^2 is known or unknown.

Acknowledgement : Authors are thankful to Prof. Jokhan Singh for his guidance in writing this article.

References

Pandey, M. and Singh,J. (1978) : A note on testing mean of a normal population when variance is unknown. J.I.S.A. 16, pp. 145-146.

Pradhan,M. (1968) : On testing a simple hypothesis against a simple alternative making the sum of the probabilities of two errors minimum. J.I.S.A. 6, pp. 149-159.

Rao, C.R.(1973) : Linear Statistical Inference and its Applications. John Wiley and Sons

Singh,J. (1985): Optimum Statistical Test Procedures. Vikram Mathematical Journal Vol V, pp. 53-56.

A Note on Testing of Hypothesis

Rajesh Singh

School of Statistics, D. A.V.V., Indore (M.P.), India

rsinghstat@yahoo.co.in

Jayant Singh

Department of Statistics

Rajasthan University, Jaipur, India

Jayantsingh47@rediffmail.com

Florentin Smarandache

Chair of Department of Mathematics, University of New Mexico, Gallup, USA

fsmarandache@yahoo.com

Abstract :

In testing of hypothesis situation if the null hypothesis is rejected will it automatically imply alternative hypothesis will be accepted. This problem has been discussed by taking examples from normal distribution.

Keywords : Hypothesis, level of significance, Baye's rule.

1. Introduction

Let the random variable (r.v.) X have a normal distribution $N(\theta, \sigma^2)$, σ^2 is assumed to be known. The hypothesis $H_0 : \theta = \theta_0$ against $H_1 : \theta = \theta_1$, $\theta_1 > \theta_0$ is to be tested. Let X_1, X_2, …, X_n be a random sample from $N(\theta, \sigma^2)$ population. Let $\overline{X} (= \frac{1}{n}\sum_{i=1}^{n} X_i)$ be the sample mean.

By Neyman – Pearson lemma the most powerful test rejects H_0 at α% level of significance,

if $\frac{\sqrt{n}\left(\overline{X} - \theta_o\right)}{\sigma} \geq d_\alpha$, where d_α is such that

$$\int_{d_\alpha}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{Z^2}{2}} dZ = \alpha$$

If the sample is such that H_0 is rejected then will it imply that H_1 will be accepted?

In general this will not be true for all values of θ_1, but will be true for some specific value of θ_1 i.e., when θ_1 is at a specific distance from θ_0.

H_0 is rejected if $\frac{\sqrt{n}\left(\overline{X} - \theta_o\right)}{\sigma} \geq d_\alpha$

$$\text{i.e. } \overline{X} \geq \theta_0 + d_\alpha \frac{\sigma}{\sqrt{n}} \qquad (1)$$

Similarly the Most Powerful Test will accept H_1 against H_0

if $\frac{\sqrt{n}\left(\overline{X} - \theta_1\right)}{\sigma} \geq -d_\alpha$

i.e. $\quad \overline{X} \geq \theta_1 - d_\alpha \frac{\sigma}{\sqrt{n}} \qquad (2)$

Rejecting H_0 will mean accepting H_1

if $\quad (1) \Rightarrow (2)$

i.e. $\quad \overline{X} \geq \theta_0 + d_\alpha \frac{\sigma}{\sqrt{n}} \quad \Rightarrow \quad \overline{X} \geq \theta_1 - d_\alpha \frac{\sigma}{\sqrt{n}}$

i.e. $\quad \theta_1 - d_\alpha \frac{\sigma}{\sqrt{n}} \leq \theta_0 + d_\alpha \frac{\sigma}{\sqrt{n}} \qquad (3)$

Similarly accepting H_1 will mean rejecting H_0

if $\quad (2) \Rightarrow (1)$

i.e. $\quad \theta_0 + d_\alpha \frac{\sigma}{\sqrt{n}} \leq \theta_1 - d_\alpha \frac{\sigma}{\sqrt{n}} \qquad (4)$

From (3) and (4) we have

$$\theta_0 + d_\alpha \frac{\sigma}{\sqrt{n}} = \theta_1 - d_\alpha \frac{\sigma}{\sqrt{n}}$$

i.e. $\quad \theta_1 - \theta_0 = 2\, d_\alpha \frac{\sigma}{\sqrt{n}} \qquad (5)$

Thus $d_\alpha \frac{\sigma}{\sqrt{n}} = \frac{\theta_1 - \theta_0}{2}$ and $\theta_1 = \theta_0 + 2\, d_\alpha \frac{\sigma}{\sqrt{n}}$.

From (1) $\quad$ Reject H_0 if $\quad \overline{X} > \theta_0 + \frac{\theta_1 - \theta_0}{2} = \frac{\theta_0 + \theta_1}{2}$

and from (2) $\;$ Accept H_1 if $\quad \overline{X} > \theta_1 - \frac{\theta_1 - \theta_0}{2} = \frac{\theta_0 + \theta_1}{2}$

Thus rejecting H_0 will mean accepting H_1

when $\quad \overline{X} > \dfrac{\theta_0 + \theta_1}{2}$.

From (5) this will be true only when $\theta_1 = \theta_0 + 2\, d_\alpha \dfrac{\sigma}{\sqrt{n}}$. For other values of

$\theta_1 \neq \theta_0 + 2\, d_\alpha \dfrac{\sigma}{\sqrt{n}}$ rejecting H_0 will not mean accepting H_1.

It is therefore, recommended that instead of testing $H_0 : \theta = \theta_0$ against $H_1 : \theta = \theta_1,\ \theta_1 > \theta_0$, it is more appropriate to test $H_0 : \theta = \theta_0$ against $H_1 : \theta > \theta_0$. In this situation rejecting H_0 will mean $\theta > \theta_0$ and is not equal to some given value θ_1.

But in Baye's setup rejecting H_0 means accepting H_1 whatever may be θ_0 and θ_1. In this set up the level of significance is not a preassigned constant, but depends on $\theta_0,\ \theta_1,\ \sigma^2$ and n.

Consider (0,1) loss function and equal prior probabilities ½ for θ_0 and θ_1. The Baye's test rejects H_0 (accepts H_1)

if $\quad \overline{X} > \dfrac{\theta_0 + \theta_1}{2}$

and accepts H_0 (rejects H_1)

if $\quad \overline{X} < \dfrac{\theta_0 + \theta_1}{2}$.

[See Rohatagi, p.463, Example 2.]

The level of significance is given by

$$P_{H_0}\left[\overline{X} > \frac{\theta_0 + \theta_1}{2}\right] = P_{H_0}\left[\frac{(\overline{X} - \theta_0)\sqrt{n}}{\sigma} > \frac{(\theta_1 - \theta_0)\sqrt{n}}{2\sigma}\right]$$

$$= 1 - \Phi\left(\frac{\sqrt{n}(\theta_1 - \theta_0)}{2\sigma}\right)$$

where $\Phi(t) = \int_{-\infty}^{t} \frac{1}{\sqrt{2\pi}} e^{-\frac{Z^2}{2}} dZ$.

Thus the level of significance depends on θ_0, θ_1, σ^2 and n.

Acknowledgement : Author's are thankful to Prof. Jokhan Singh for suggesting this problem.

References

Lehmann, E.L. (1976) : Testing Statistical Hypotheses, Wiley Eastern Ltd., New Delhi.

Rohatagi, V.K. (1985) : An introduction to probability theory and mathematical statistics, Wiley Eastern Ltd., New Delhi.

Improvement in Estimating Population Mean using Two Auxiliary Variables in Two-Phase Sampling

Rajesh Singh
Department of Statistics, Banaras Hindu University(U.P.), India
(rsinghstat@yahoo.com)

Pankaj Chauhan and Nirmala Sawan
School of Statistics, DAVV, Indore (M.P.), India

Florentin Smarandache
Department of Mathematics, University of New Mexico, Gallup, USA
(smarand@unm.edu)

Abstract

This study proposes improved chain-ratio type estimator for estimating population mean using some known values of population parameter(s) of the second auxiliary character. The proposed estimators have been compared with two-phase ratio estimator and some other chain ratio type estimators. The performances of the proposed estimators have been supported with a numerical illustration.

Key words: Auxiliary variables, chain ratio-type estimator, bias, mean squared error.

1. Introduction

The ratio method of estimation is generally used when the study variable Y is positively correlated with an auxiliary variable X whose population mean is known in advance. In the absence of the knowledge on the population mean of the auxiliary character we go for two-phase (double) sampling. The two-phase sampling happens to be a powerful and cost effective (economical) procedure for finding the reliable estimate in first phase sample for the unknown parameters of the auxiliary variable x and hence has eminent role to play in survey sampling, for instance, see Hidiroglou and Sarndal (1998).

Consider a finite population $U = (U_1, U_2, ..., U_N)$. Let y and x be the study and auxiliary variable, taking values y_i and x_i respectively for the i^{th} unit U_i.

Allowing SRSWOR (Simple Random Sampling without Replacement) design in each phase, the two-phase sampling scheme is as follows:

(i) the first phase sample $s_{n'}$ $(s_{n'} \subset U)$ of a fixed size n' is drawn to measure only x in order to formulate a good estimate of a population mean $\overline{X}$,

(ii) Given $s_{n'}$, the second phase sample s_n $(s_n \subset s_{n'})$ of a fixed size n is drawn to measure y only.

Let $\overline{x} = \frac{1}{n}\sum_{i \in s_n} x_i$, $\overline{y} = \frac{1}{n}\sum_{i \in s_n} y_i$ and $\overline{x}' = \frac{1}{n'}\sum_{i \in s_{n'}} x_i$.

The classical ratio estimator for $\overline{Y}$ is defined as

$$\overline{y}_r = \frac{\overline{y}}{\overline{x}}\overline{X} \tag{1.1}$$

If $\overline{X}$ is not known, we estimate $\overline{Y}$ by two-phase ratio estimator

$$\overline{y}_{rd} = \frac{\overline{y}}{\overline{x}}\overline{x}' \tag{1.2}$$

Some times even if $\overline{X}$ is not known, information on a cheaply ascertainable variable z, closely related to x but compared to x remotely related to y, is available on all units of the population. For instance, while estimating the total yield of wheat in a village, the yield and area under the crop are likely to be unknown, but the total area of each farm may be known from village records or may be obtained at a low cost. Then y, x and z are respectively yield, area under wheat and area under cultivation see Singh et al. (2004).

Assuming that the population mean $\overline{Z}$ of the variable z is known, Chand (1975) proposed a chain type ratio estimator as

$$t_1 = \frac{\overline{y}}{\overline{x}}\left(\frac{\overline{x}'}{\overline{z}'}\right)\overline{Z} \tag{1.3}$$

Several authors have used prior value of certain population parameter(s) to find more precise estimates. Singh and Upadhyaya (1995) used coefficient of variation of z for defining

modified chain type ratio estimator. In many situation the value of the auxiliary variable may be available for each unit in the population, for instance, see Das and Tripathi (1981). In such situations knowledge on $\bar{Z}$, C_z, $\beta_1(z)$ (coefficient of skewness), $\beta_2(z)$ (coefficient of kurtosis) and possibly on some other parameters may be utilized. Regarding the availability of information on C_z, $\beta_1(z)$ and $\beta_2(z)$, the researchers may be referred to Searls(1964), Sen(1978), Singh et al.(1973), Searls and Intarapanich(1990) and Singh et.al.(2007). Using the known coefficient of variation C_z and known coefficient of kurtosis $\beta_2(z)$ of the second auxiliary character z Upadhyaya and Singh (2001) proposed some estimators for $\bar{Y}$.

If the population mean and coefficient of variation of the second auxiliary character is known, the standard deviation σ_z is automatically known and it is more meaningful to use the σ_z in addition to C_z, see Srivastava and Jhajj (1980). Further, C_z, $\beta_1(z)$ and $\beta_2(z)$ are the unit free constants, their use in additive form is not much justified. Motivated with the above justifications and utilizing the known values of σ_z, $\beta_1(z)$ and $\beta_2(z)$, Singh (2001) suggested some modified estimators for $\bar{Y}$.

In this paper, under simple random sampling without replacement (SRSWOR), we have suggested improved chain ratio type estimator for estimating population mean using some known values of population parameter(s).

2. The suggested estimator

The work of authors discussed in section 1 can be summarized by using following estimator

$$t = \bar{y}\left(\frac{\bar{x}'}{\bar{x}}\right)\left(\frac{a\bar{Z}+b}{a\bar{z}'+b}\right) \tag{2.1}$$

where $a(\neq 0)$, b are either real numbers or the functions of the known parameters of the second auxiliary variable z such as standard deviation (σ_z), coefficient of variation (C_z), skewness ($\beta_1(z)$) and kurtosis ($\beta_2(z)$).

The following scheme presents some of the important known estimators of the population mean which can be obtained by suitable choice of constants a and b.

Estimator	Values of	
	a	b
$t_1 = \bar{y}\left(\frac{\bar{x}'}{\bar{x}}\right)\left(\frac{\bar{Z}}{\bar{z}'}\right)$ Chand (1975) chain ratio type estimator	1	0
$t_2 = \bar{y}\left(\frac{\bar{x}'}{\bar{x}}\right)\left(\frac{\bar{Z}+C_z}{\bar{z}'+C_z}\right)$ Singh and Upadhyaya (1995) estimator	1	C_z
$t_3 = \bar{y}\left(\frac{\bar{x}'}{\bar{x}}\right)\left(\frac{\beta_2(z)\bar{Z}+C_z}{\beta_2(z)\bar{z}'+C_z}\right)$ Upadhyaya and Singh (2001) estimator	$\beta_2(z)$	C_z
$t_4 = \bar{y}\left(\frac{\bar{x}'}{\bar{x}}\right)\left(\frac{C_z\bar{Z}+\beta_2(z)}{C_z\bar{z}'+\beta_2(z)}\right)$ Upadhyaya and Singh (2001) estimator	C_z	$\beta_2(z)$
$t_5 = \bar{y}\left(\frac{\bar{x}'}{\bar{x}}\right)\left(\frac{\bar{Z}+\sigma_z}{\bar{z}'+\sigma_z}\right)$ Singh (2001) estimator	1	σ_z
$t_6 = \bar{y}\left(\frac{\bar{x}'}{\bar{x}}\right)\left(\frac{\beta_1(z)\bar{Z}+\sigma_z}{\beta_1(z)\bar{z}'+\sigma_z}\right)$ Singh (2001) estimator	$\beta_1(z)$	σ_z
$t_7 = \bar{y}\left(\frac{\bar{x}'}{\bar{x}}\right)\left(\frac{\beta_2(z)\bar{Z}+\sigma_z}{\beta_2(z)\bar{z}'+\sigma_z}\right)$	$\beta_2(z)$	σ_z

In addition to these estimators a large number of estimators can also be generated from the estimator t at (2.1) by putting suitable values of a and b.

Following Kadilar and Cingi (2006), we propose modified estimator combining t_1 and t_i $(i = 2,3,....,7)$ as follows

$$t_i^* = \alpha t_1 + (1-\alpha)t_i\,,\ (i = 2,3,....,7) \tag{2.2}$$

where α is a real constant to be determined such that MSE of t_i^* is minimum and t_i $(i = 2,3,....,7)$ are estimators listed above.

To obtain the bias and MSE of t_i^*, we write

$$\bar{y} = \bar{Y}(1+e_0),\ \bar{x} = \bar{X}(1+e_1),\ \bar{x}' = \bar{X}(1+e_1'),\ \bar{z}' = \bar{Z}(1+e_2')$$

such that

$$E(e_0) = E(e_1) = E(e_1') = E(e_2') = 0$$

and

$$E(e_0^2) = f_1 C_y^2,\qquad E(e_1^2) = f_1 C_x^2,\qquad E(e_1'^2) = f_2 C_x^2$$

$$E(e_2'^2) = f_2 C_z^2,\qquad E(e_0 e_1) = f_1 \rho_{xy} C_x C_y,\quad E(e_0 e_1') = f_2 \rho_{xy} C_x C_y$$

$$E(e_0 e_2') = f_2 \rho_{yz} C_y C_z,\quad E(e_1 e_1') = f_2 C_x^2,\qquad E(e_1 e_2') = f_2 \rho_{xz} C_x C_z$$

$$E(e_1' e_2') = f_2 \rho_{xz} C_x C_z$$

where

$$f_1 = \left(\frac{1}{n} - \frac{1}{N}\right),\qquad f_2 = \left(\frac{1}{n'} - \frac{1}{N}\right),$$

$$C_y^2 = \frac{S_y^2}{\bar{Y}^2},\qquad C_x^2 = \frac{S_x^2}{\bar{X}^2},\qquad C_z^2 = \frac{S_z^2}{\bar{Z}^2}$$

$$\rho_{xy} = \frac{S_{xy}}{S_x S_y},\qquad \rho_{xz} = \frac{S_{xz}}{S_x S_z},\qquad \rho_{yz} = \frac{S_{yz}}{S_y S_z}$$

$$S_y^2 = \frac{1}{(N-1)}\sum_{i\in U}(y_i - \bar{Y})^2,\qquad S_x^2 = \frac{1}{(N-1)}\sum_{i\in U}(x_i - \bar{X})^2$$

$$S_z^2 = \frac{1}{(N-1)}\sum_{i\in U}(z_i - \bar{Z})^2,\qquad S_{xy} = \frac{1}{(N-1)}\sum_{i\in U}(x_i - \bar{X})(y_i - \bar{Y})$$

$$S_{xz}=\frac{1}{(N-1)}\sum_{i\in U}\left(x_i-\overline{X}\right)\left(z_i-\overline{Z}\right),\qquad S_{yz}=\frac{1}{(N-1)}\sum_{i\in U}\left(y_i-\overline{Y}\right)\left(z_i-\overline{Z}\right).$$

Expressing t_i^* in terms of e's, we have

$$t_i^*=\overline{Y}(1+e_0)\left[\alpha(1+e_1')(1+e_1)^{-1}(1+e_2')^{-1}+(1-\alpha)(1+e_1')(1+e_1)^{-1}(1+\theta e_2')^{-1}\right] \tag{2.3}$$

where $\theta=\dfrac{a\overline{Z}}{a\overline{Z}+b}$ (2.4)

Expanding the right hand side of (2.3) and retaining terms up to second power of e's, we have

$$t_i^*\cong\overline{Y}\left[1+e_0-e_1+e_1'-e_2'(\alpha+\theta-\alpha\theta)\right] \tag{2.5}$$

or

$$t_i^*-\overline{Y}\cong\overline{Y}\left[e_0-e_1+e_1'-e_2'(\alpha+\theta-\alpha\theta)\right] \tag{2.6}$$

Squaring both sides of (2.6) and then taking expectation, we get the MSE of the estimator t_i^*, up to the first order of approximation, as

$$MSE(t_i^*)=\overline{Y}^2\left[f_1C_y^2+f_3C_x^2+(\alpha+\theta-\alpha\theta)^2f_2C_z^2-2f_3\rho C_yC_x-2(\alpha+\theta-\alpha\theta)f_2\rho C_yC_z\right]$$

(2.7)

where

$$f_3=\left(\frac{1}{n}-\frac{1}{n'}\right).$$

Minimization of (2.7) with respect to α yield its optimum value as

$$\alpha_{opt}=\frac{K_{yz}-\theta}{1-\theta} \tag{2.8}$$

where

$$K_{yz}=\rho_{yz}\frac{C_y}{C_z}.$$

Substitution of (2.8) in (2.7) yields the minimum value of MSE (t_i^*) as –

$$\min.\text{MSE}(t_i^*) = M_o = \overline{Y}^2\left[f_1C_y^2 + f_3(C_x^2 - 2\rho_{yx}C_yC_x) - f_2\rho_{yz}^2C_y^2\right] \tag{2.9}$$

3. Efficiency comparisons

In this section, the conditions for which the proposed estimator is better than $t_i (i = 1,2,....7)$ have been obtained. The MSE's of these estimators up to the order $o(n)^{-1}$ are derived as –

$$\text{MSE}(\overline{y}_{rd}) = \overline{Y}^2\left[f_1C_y^2 + f_3(C_x^2 - 2\rho_{yx}C_yC_x)\right] \tag{3.1}$$

$$\text{MSE}(t_1) = \overline{Y}^2\left[f_1C_y^2 + f_2(C_z^2 - 2\rho_{yz}C_yC_z) + f_3(C_x^2 - 2\rho_{yx}C_yC_x)\right] \tag{3.2}$$

$$\text{MSE}(t_2) = \overline{Y}^2\left[f_1C_y^2 + f_2(\theta_2^2C_z^2 - 2\theta_2\rho_{yz}C_yC_z) + f_3(C_x^2 - 2\rho_{yx}C_yCx)\right] \tag{3.3}$$

$$\text{MSE}(t_3) = \overline{Y}^2\left[f_1C_y^2 + f_2(\theta_3^2C_z^2 - 2\theta_3\rho_{yz}C_yC_z) + f_3(C_x^2 - 2\rho_{yx}C_yCx)\right] \tag{3.4}$$

$$\text{MSE}(t_4) = \overline{Y}^2\left[f_1C_y^2 + f_2(\theta_4^2C_z^2 - 2\theta_4\rho_{yz}C_yC_z) + f_3(C_x^2 - 2\rho_{yx}C_yCx)\right] \tag{3.5}$$

$$\text{MSE}(t_5) = \overline{Y}^2\left[f_1C_y^2 + f_2(\theta_5^2C_z^2 - 2\theta_5\rho_{yz}C_yC_z) + f_3(C_x^2 - 2\rho_{yx}C_yCx)\right] \tag{3.6}$$

$$\text{MSE}(t_6) = \overline{Y}^2\left[f_1C_y^2 + f_2(\theta_6^2C_z^2 - 2\theta_6\rho_{yz}C_yC_z) + f_3(C_x^2 - 2\rho_{yx}C_yCx)\right] \tag{3.7}$$

and

$$\text{MSE}(t_7) = \overline{Y}^2\left[f_1C_y^2 + f_2(\theta_7^2C_z^2 - 2\theta_7\rho_{yz}C_yC_z) + f_3(C_x^2 - 2\rho_{yx}C_yCx)\right] \tag{3.8}$$

where

$$\theta_2 = \frac{\overline{Z}}{\overline{Z} + C_z},\ \theta_3 = \frac{\beta_2(z)\overline{Z}}{\beta_2(z)\overline{Z} + C_z},\ \theta_4 = \frac{C_z\overline{Z}}{C_2\overline{Z} + \beta_2(z)},\ \theta_5 = \frac{\overline{Z}}{\overline{Z} + \sigma_z},$$

$$\theta_6 = \frac{\beta_1(z)\overline{Z}}{\beta_1(z)\overline{Z} + \sigma_z},\ \theta_7 = \frac{\beta_2(z)\overline{Z}}{\beta_2(z)\overline{Z} + \sigma_z}.$$

From (2.9) and (3.1), we have

$$\text{MSE}(\overline{y}_{rd}) - M_o = f_2\rho_{yz}^2C_y^2 \geq 0 \tag{3.9}$$

Also from (2.9) and (3.2)-(3.8), we have

$$MSE(t_i) - M_o = f_2\left(\theta_i C_z - \rho_{yz} C_y\right)^2 \geq 0, \; \left(i = 2,3,....,7\right) \qquad (3.10)$$

Thus it follows from (3.9) and (3.10) that the suggested estimator under optimum condition is always better than the estimator $t_i (i = 1,2,....7)$.

4. Empirical study

To illustrate the performance of various estimators of $\overline{Y}$, we consider the data used by Anderson (1958). The variates are

y : Head length of second son

x : Head length of first son

z : Head breadth of first son

$N = 25$, $\overline{Y} = 183.84 =$, $\overline{X} = 185.72$, $\overline{Z} = 151.12$, $\sigma_z = 7.224$, $C_y = 0.0546$, $C_x = 0.0526$, $C_z = 0.0488$, $\rho_{yx} = 0.7108$, $\rho_{yz} = 0.6932$, $\rho_{xz} = 0.7346$, $\beta_1(z) = 0.002$, $\beta_2(z) = 2.6519$.

Consider $n' = 10$ and n = 7.

We have computed the percent relative efficiency (PRE) of different estimators of $\overline{Y}$ with respect to usual estimator $\overline{y}$ and compiled in the table 4.1:

Table 4.1: PRE of different estimators of $\overline{Y}$ with respect to $\overline{y}$

estimator	PRE
$\overline{y}$	100
$\overline{y}_{rd}$	122.5393
t_1	178.8189
t_2	178.8405
t_3	178.8277
t_4	186.3912
t_5	181.6025
t_6	122.5473

t_7	179.9636
t_i^*	186.6515

5. Conclusion

We have suggested modified estimators $t_i^*(i = 2,3,....,7)$. From table 4.1, we conclude that the proposed estimators are better than usual two-phase ratio estimator $\overline{y}_{rd}$, Chand (1975) chain type ratio estimator t_1, estimator t_2 proposed by Singh and Upadhyaya (1995), estimators $t_i(i = 3,4)$ and than that of Singh (2001) estimators $t_i(i = 5,6,7)$. For practical purposes the choice of the estimator depends upon the availability of the population parameter(s).

References

Anderson, T. W. (1958), An Introduction to Multivariate Statistical Analysis. John Wiley & Sons, Inc., New York.

Chand, L. (1975): Some ratio type estimators based on two or more auxiliary variables. Unpublished Ph. D. thesis, Iowa State University, Ames, Iowa (USA).

Das, A. K. and Tripathi, T. P. (1981): a class of sampling strategies for population mean using information on mean and variance of an auxiliary character. Proc. of the Indian Statistical Institute Golden Jubilee International Conference on Statistics. Applications and New Directions, Calcutta, 16-19, December 1981, 174-181.

Hidiroglou, M. A. and Sarndal, C.E. (1998): Use of auxiliary information for two-phase sampling. Survey Methodology, 24(1), 11-20.

Searls, D.T. (1964): The utilization of known coefficient of variation in the estimation procedure. Journal of American Statistical Association, 59, 1125-1126.

Searls, D.T. and Intarapanich, R. (1990): A note on an estimator for variance that utilizes the kurtosis. Amer. Stat., 44(4), 295-296.

Sen, A.R. (1978): Estimation of the population mean when the coefficient of variation is known. Comm. Stat.-Theory Methods, A7, 657-672.

Singh, G. N. (2001): On the use of transformed auxiliary variable in the estimation of population mean in two-phase sampling. Statistics in Transition, 5(3), 405-416.

Singh, G. N. and Upadhyaya, L. N. (1995): A class of modified chain type estimators using two auxiliary variables in two-phase sampling. Metron, LIII, 117-125.

Singh, H. P, Upadhyaya, L. N. and Chandra, P. (2004): A general family of estimators for estimating population mean using two auxiliary variables in two-phase sampling. Statistics in transition, 6(7), 1055-1077.

Singh, J., Pandey, B. N. and Hirano, K. (1973): On the utilization of a known coefficient of kurtosis in the estimation procedure of variance. Ann. Inst. Statist. Math., 25, 51-55.

Singh, R. Chauhan, P. Sawan, N. and Smarandache, F. (2007): Auxiliary information and a priori values in construction of improved estimators. Renaissance high press, USA.

Srivastava, S.K. and Jhajj, H.S. (1980): A class of estimators using auxiliary information for estimating finite population variance. Sankhya, C, 42, 87-96.

Upadhyaya, L. N. and Singh, G. N. (2001): Chain type estimators using transformed auxiliary variable in two-phase sampling. Advances in Modeling and Analysis, 38, (1-2), 1-10.

Improved Exponential Estimator for Population Variance Using Two Auxiliary Variables

Rajesh Singh

Department of Statistics,Banaras Hindu University(U.P.), India

(rsinghstat@yahoo.com)

Pankaj Chauhan and Nirmala Sawan

School of Statistics, DAVV, Indore (M.P.), India

Florentin Smarandache

Department of Mathematics, University of New Mexico, Gallup, USA

(smarand@unm.edu)

Abstract

In this paper exponential ratio and exponential product type estimators using two auxiliary variables are proposed for estimating unknown population variance S_y^2 . Problem is extended to the case of two-phase sampling. Theoretical results are supported by an empirical study.

Key words: Auxiliary information, exponential estimator, mean squared error.

1. Introduction

It is common practice to use the auxiliary variable for improving the precision of the estimate of a parameter. Out of many ratio and product methods of estimation are good examples in this context. When the correlation between the study variate and the auxiliary variate is positive (high) ratio method of estimation is quite effective. On the other hand, when this correlation is negative (high) product method of estimation can be employed effectively. Let y and (x,z) denotes the study variate and auxiliary variates taking the values y_i and (x_i,z_i) respectively, on the unit U_i (i=1,2,……,N), where x is positively correlated with

y and z is negatively correlated with y. To estimate $S_y^2 = \frac{1}{(N-1)}\sum_{i=1}^{N}(y_i - \bar{y})^2$, it is assumed that $S_x^2 = \frac{1}{(N-1)}\sum_{i=1}^{N}(x_i - \bar{X})^2$ and $S_z^2 = \frac{1}{(N-1)}\sum_{i=1}^{N}(z_i - \bar{Z})^2$ are known. Assume that population size N is large so that the finite population correction terms are ignored.

Assume that a simple random sample of size n is drawn without replacement (SRSWOR) from U. The usual unbiased estimator of S_y^2 is

$$s_y^2 = \frac{1}{(n-1)}\sum_{i=1}^{n}(y_i - \bar{y})^2 \tag{1.1}$$

where $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$ is the sample mean of y.

When the population variance $S_x^2 = \frac{1}{(N-1)}\sum_{i=1}^{N}(x_i - \bar{X})^2$ is known, Isaki (1983) proposed a ratio estimator for S_y^2 as

$$t_k = s_y^2 \frac{S_x^2}{s_x^2} \tag{1.2}$$

where $s_x^2 = \frac{1}{(n-1)}\sum_{i=1}^{n}(x_i - \bar{X})^2$ is an unbiased estimator of S_x^2.

Up to the first order of approximation, the variance of S_y^2 and MSE of t_k (ignoring the finite population correction (fpc) term) are respectively given by

$$\operatorname{var}\left(s_y^2\right) = \left(\frac{S_y^4}{n}\right)\left[\partial_{400} - 1\right] \tag{1.3}$$

$$\operatorname{MSE}\left(t_k\right) = \left(\frac{S_y^4}{n}\right)\left[\partial_{400} + \partial_{040} - 2\partial_{220}\right] \tag{1.4}$$

where $\delta_{pqr} = \dfrac{\mu_{pqr}}{\left(\mu_{200}^{p/2}\mu_{020}^{q/2}\mu_{002}^{r/2}\right)}$,

$\mu_{pqr} = \dfrac{1}{N}\sum_{i=1}^{N}\left(y_i - \overline{Y}\right)^p\left(x_i - \overline{X}\right)^q\left(z_i - \overline{Z}\right)^r$; p, q, r being the non-negative integers.

Following Bahl and Tuteja (1991), we propose exponential ratio type and exponential product type estimators for estimating population variance S_y^2 as –

$$t_1 = s_y^2 \exp\left[\frac{S_x^2 - s_x^2}{S_x^2 + s_x^2}\right] \tag{1.5}$$

$$t_2 = s_y^2 \exp\left[\frac{s_z^2 - S_z^2}{s_z^2 + S_z^2}\right] \tag{1.6}$$

2. **Bias and MSE of proposed estimators**

To obtain the bias and MSE of t_1, we write

$$s_y^2 = S_y^2\left(1 + e_0\right),\ s_x^2 = S_x^2\left(1 + e_1\right)$$

Such that $E(e_0) = E(e_1) = 0$

and $E(e_0^2) = \dfrac{1}{n}\left(\partial_{400} - 1\right),\ E(e_1^2) = \dfrac{1}{n}\left(\partial_{040} - 1\right),\ E(e_0 e_1) = \dfrac{1}{n}\left(\partial_{220} - 1\right).$

After simplification we get the bias and MSE of t_1 as

$$B(t_1) \cong \frac{S_y^2}{n}\left[\frac{\partial_{040}}{8} - \frac{\partial_{220}}{2} + \frac{3}{8}\right] \tag{2.1}$$

$$MSE(t_1) \cong \frac{S_y^2}{n}\left[\partial_{400} + \frac{\partial_{040}}{4} - \partial_{220} + \frac{1}{4}\right] \tag{2.2}$$

To obtain the bias and MSE of t_2, we write

$$s_y^2 = S_y^2\left(1 + e_0\right),\ s_z^2 = S_z^2\left(1 + e_2\right)$$

Such that $E(e_0) = E(e_2) = 0$

$$E(e_2^2)=\frac{1}{n}(\partial_{004}-1),\ E(e_0e_2)=\frac{1}{n}(\partial_{202}-1)$$

After simplification we get the bias and MSE of t_2 as

$$B(t_2)\cong\frac{S_y^2}{n}\left[\frac{\partial_{004}}{8}+\frac{\partial_{202}}{2}-\frac{5}{8}\right] \tag{2.3}$$

$$MSE(t_2)\cong\frac{S_y^2}{n}\left[\partial_{400}+\frac{\partial_{004}}{4}+\partial_{202}-\frac{9}{4}\right] \tag{2.4}$$

3. Improved Estimator

Following Kadilar and Cingi (2006) and Singh et. al. (2007), we propose an improved estimator for estimating population variance S_y^2 as-

$$t=s_y^2\left[\alpha\exp\left\{\frac{S_x^2-s_x^2}{S_x^2+s_x^2}\right\}+(1-\alpha)\exp\left\{\frac{s_z^2-S_z^2}{s_x^2+S_z^2}\right\}\right] \tag{3.1}$$

where α is a real constant to be determined such that the MSE of t is minimum.

Expressing t in terms of e's, we have

$$t=S_y^2(1+e_0)\left[\alpha\exp\left\{-\frac{e_1}{2}\left(1+\frac{e_1}{2}\right)^{-1}\right\}+(1-\alpha)\exp\left\{\frac{e_2}{2}\left(1+\frac{e_2}{2}\right)^{-1}\right\}\right] \tag{3.2}$$

Expanding the right hand side of (3.2) and retaining terms up to second power of e's, we have

$$t\cong S_y^2\left[1+e_0+\frac{e_2}{2}+\frac{e_2^2}{8}+\frac{e_0e_2}{2}+\alpha\left(-\frac{e_1}{2}+\frac{e_1^2}{8}\right)-\alpha\left(\frac{e_2}{2}+\frac{e_2^2}{8}\right)\right.$$

$$\left.+e_0\alpha\left(-\frac{e_1}{2}+\frac{e_1^2}{8}\right)-\alpha e_0\left(\frac{e_2}{2}+\frac{e_2^2}{8}\right)\right] \tag{3.3}$$

Taking expectations of both sides of (3.3) and then subtracting S_y^2 from both sides, we get the bias of the estimator t, up to the first order of approximation, as

$$B(t)=\frac{S_y^2}{n}\left[\frac{\alpha}{8}(\partial_{040}-1)+\frac{(1-\alpha)}{8}(\partial_{004}-1)+\frac{(1-\alpha)}{2}(\partial_{202}-1)\right.$$

$$\left.-\frac{\alpha}{2}(\partial_{220}-1)\right] \tag{3.4}$$

From (3.4), we have

$$\left(t-S_y^2\right)\cong S_y^2\left[e_0-\frac{\alpha e_1}{2}+\frac{(1-\alpha)}{2}e_2\right] \tag{3.5}$$

Squaring both the sides of (3.5) and then taking expectation, we get MSE of the estimator t, up to the first order of approximation, as

$$\mathrm{MSE}(t)\cong\frac{S_y^4}{n}\left[(\partial_{400}-1)+\frac{\alpha^2}{4}(\partial_{040}-1)+\frac{(1-\alpha^2)}{4}(\partial_{004}-1)\right.$$

$$\left.-\alpha(\partial_{220}-1)+(1-\alpha)(\partial_{202}-1)-\frac{\alpha(1-\alpha)}{2}(\partial_{022}-1)\right] \tag{3.6}$$

Minimization of (3.6) with respect to α yields its optimum value as

$$\alpha=\frac{\{\partial_{004}+2(\partial_{220}+\partial_{202})+\partial_{022}-6\}}{(\partial_{040}+\partial_{004}+2\partial_{022}-4)}=\alpha_0\,(\text{say}) \tag{3.7}$$

Substitution of α_0 from (3.7) into (3.6) gives minimum value of MSE of t.

4. Proposed estimators in two-phase sampling

In certain practical situations when S_x^2 is not known a priori, the technique of two-phase or double sampling is used. This scheme requires collection of information on x and z the first phase sample s' of size n' (n'<N) and on y for the second phase sample s of size n (n<n') from the first phase sample.

The estimators t_1, t_2 and t in two-phase sampling will take the following form, respectively

$$t_{1d}=s_y^2\exp\left[\frac{s'^2_x-s_x^2}{s'^2_x+s_x^2}\right] \tag{4.1}$$

$$t_{2d} = s_z^2 \exp\left[\frac{s'^2_z - s_z^2}{s'^2_z + s_z^2}\right] \tag{4.2}$$

$$t_d = s_y^2\left[k \exp\left\{\frac{s'^2_x - s_x^2}{s'^2_x + s_x^2}\right\} + (1-k)\exp\left\{\frac{s'^2_z - s_z^2}{s'^2_z + s_z^2}\right\}\right] \tag{4.3}$$

To obtain the bias and MSE of t_{1d}, t_{2d}, t_d, we write

$$s_y^2 = S_y^2(1+e_0),\ s_x^2 = S_x^2(1+e_1),\ s'^2_x = S_x^2(1+e'_1)$$

$$s_z^2 = S_z^2(1+e_2),\ s'^2_z = S_z^2(1+e'_2)$$

where $s'^2_x = \frac{1}{(n'-1)}\sum_{i=1}^{n'}(x_i - \bar{x}')^2$, $s_z^2 = \frac{1}{(n'-1)}\sum_{i=1}^{n'}(z_i - \bar{z}')^2$

$$\bar{x}' = \frac{1}{n'}\sum_{i=1}^{n'} x_i\ ,\ \bar{z}' = \frac{1}{n'}\sum_{i=1}^{n'} z_i$$

Also,

$E(e'_1) = E(e'_2) = 0$,

$$E(e'^2_1) = \frac{1}{n'}(\partial_{040} - 1),\ E(e'^2_2) = \frac{1}{n}(\partial_{004} - 1),$$

$$E(e'_1\, e'_2) = \frac{1}{n'}(\partial_{220} - 1)$$

Expressing t_{1d}, t_{2d}, and t_d in terms of e's and following the procedure explained in section 2 and section3 we get the MSE of these estimators, respectively as-

$$MSE(t_{1d}) \cong S_y^4\left[\frac{1}{n}(\partial_{400} - 1) + \frac{1}{4}\left(\frac{1}{n} - \frac{1}{n'}\right)(\partial_{040} - 1)\right.$$
$$\left. + \left(\frac{1}{n'} - \frac{1}{n}\right)(\partial_{220} - 1)\right] \tag{4.4}$$

$$\mathrm{MSE}(t_{2d}) \cong S_y^4\left[\frac{1}{n}(\partial_{400}-1)+\frac{1}{4}\left(\frac{1}{n}-\frac{1}{n'}\right)(\partial_{004}-1)\right.$$

$$\left.-\left(\frac{1}{n'}-\frac{1}{n}\right)(\partial_{202}-1)\right] \qquad (4.5)$$

$$\mathrm{MSE}(t_d) \cong S_y^4\left[\frac{1}{n}(\partial_{400}-1)+\frac{k^2}{4}\left(\frac{1}{n}-\frac{1}{n'}\right)(\partial_{040}-1)+\frac{(k^2-1)}{4}\left(\frac{1}{n}-\frac{1}{n'}\right)(\partial_{004}-1)\right.$$

$$+k\left(\frac{1}{n}-\frac{1}{n'}\right)(\partial_{220}-1)+(k-1)\left(\frac{1}{n'}-\frac{1}{n}\right)(\partial_{202}-1)$$

$$-\frac{k(k-1)}{2}\left(\frac{1}{n'}-\frac{1}{n}\right)(\partial_{022}-1) \qquad (4.6)$$

Minimization of (4.6) with respect to k yields its optimum value as

$$k=\frac{\{\partial_{004}+2(\partial_{220}-1)+\partial_{022}-6\}}{(\partial_{040}+\partial_{004}+2\partial_{022}-4)}=k_0\,(\text{say}) \qquad (4.7)$$

Substitution of k_0 from (4.7) to (4.6) gives minimum value of MSE of t_d.

5. Empirical Study

To illustrate the performance of various estimators of S_y^2, we consider the data given in Murthy(1967, p.-226). The variates are:

y: output, x: number of workers, z: fixed capital,

N=80, n'=25, n=10.

$\partial_{400}=2.2667$, $\partial_{040}=3.65$, $\partial_{004}=2.8664$, $\partial_{220}=2.3377$, $\partial_{202}=2.2208$, $\partial_{400}=3.14$

The percent relative efficiency (PRE) of various estimators of S_y^2 with respect to conventional estimator s_y^2 has been computed and displayed in table 5.1.

Table 5.1 : PRE of s_y^2, t_1, t_2 and min. MSE (t) with respect to s_y^2

Estimator	PRE(., s_y^2)
s_y^2	100
t_1	214.35
t_2	42.90
t	215.47

In table 5.2 PRE of various estimators of s_y^2 in two-phase sampling with respect to S_y^2 are displayed.

Table 5.2 : PRE of s_y^2, t_{1d}, t_{2d} and min.MSE (t_d) with respect to s_y^2

Estimator	PRE (., s_y^2)
s_y^2	100
t_{1d}	1470.76
t_{2d}	513.86
t_d	1472.77

6. Conclusion

From table 5.1 and 5.2, we infer that the proposed estimators t performs better than conventional estimator s_y^2 and other mentioned estimators.

References

Isaki, C. T. (1983): Variance estimation using auxiliary information. Journal of American Statistical Association.

Bahl, S. and Tuteja, R.K. (1991): Ratio and Product type exponential estimator, Information and Optimization sciences, Vol. XII, I, 159-163.

Kadilar,C. and Cingi,H. (2006) : Improvement in estimating the population mean in simple random sampling. Applied Mathematics Letters 19 (2006), 75–79.

Singh, R. Chauhan, P. Sawan, N. and Smarandache, F. (2007): Auxiliary information and a priori values in construction of improved estimators. Renaissance high press, USA.

Murthy, M.N.(1967): Sampling Theory and Methods. Statistical Publishing Society, Calcutta, India.

STRUCTURAL DYNAMICS OF VARIOUS CAUSES OF MIGRATION IN JAIPUR

Jayant Singh and Hansraj Yadav
Department of Statistics,
University of Rajasthan, Jaipur, India
jayantsingh47@rediffmail.com, hansraj_yadav@rediff.com
Rajesh Singh
Department of Statistics, BHU, Varanasi (U.P.), India
rsinghstat@yahoo.com
Florentin Smarandache
Chair of Department of Mathematics, University of New Mexico, Gallup, USA
fsmarandache@yahoo.com

Abstract

Jaipur urban area has grown tremendously in last three decades. Composition of People migrating due to various reasons has display a meticulous trend. Dominance of people moving due to marriages is getting sturdy whereas Jaipur city is losing its lustre in attracting persons for education and business. Short duration migration from Jaipur district to urban area has gone down to a very low level. Flow of migrants from Rural areas to Jaipur outpaced the migrants from urban areas and its composition from various in terms long and short distances migration has substantially changed over two consecutive decades. Movements of males and females were differ on many criterion as male moving faster than females for employment & education and females move faster than male for marriages and moving along family was found evident in short, medium and long distances migration. Gender gap in

people migration from different reasons was observed and a gender specific trend was seen favour. Short duration migration and migration due to education & employment is not as prominence as it was two decade back.

Introduction:

Migration from one area to another in search of improved livelihoods is a key feature of human history. While some regions and sectors fall behind in their capacity to support populations, others move ahead and people migrate to access these emerging opportunities. There are various causes like political, cultural, social, personal and natural forces but aspire for betterment, higher earning, more employment opportunities receive special attention. There are four type of migration namely

i. Rural-Rural
ii. Rural-Urban
iii. Urban-Urban
iv. Urban-Rural

Though all of these have different implication over the various demographic and socio-economic characteristics of the society but rural-urban & urban-urban migration is a cause of concern in reference to migration process to Jaipur urban agglomeration. The dynamics of migration for three census (1981, 1991, 2001) has been analysed from different angles at destination i.e. Jaipur Urban Agglomeration. The peoples of two places have different socio-economic character like education attainment, availability of land to the rural labour and agriculture production capacity,

industrialization etc and the difference of these factors at two places gear the migration process.

Distance plays a prominent role in migration of peoples, in general people from nearby area show a faster pace than the distant places due to psychic of being come back or feel like at home or the reason that some acquaintance in nearby area plays a big pull factor. However these assumptions do govern by other consideration of pull and push factor and the prevalent socio-economic aspects of the origin and destination places.

Jaipur being the capital of the state and proximity to the national state has been a great potential to draw peoples. It has not been attracting peoples from the nearby areas but it has influence on the persons of entire state and other states of the country. Majority of immigrants to Jaipur belongs to different parts of the states followed by its adjoining states. However it has been able to attract people from all over the country and overseas as well though their contribution in totality is not as significant. Seeing at this scenario it is worthwhile to limit the migrants from the following area to comprehend the migrant process of Jaipur. In-migrants to Jaipur urban area from (a) various parts of Jaipur district (b) other districts of the state (c) adjoining states of the state having fair share in migrants and (d) total migration which is overall migration from all the areas.

COMPOSITION OF IN-MIGRANTS TO JAIPUR:

In-migrants to Jaipur has grown by leaps and bounds in the last three decades. The decadal growth of in-migrants to Jaipur in last four decades synchronized with the growth of urban population of the Jaipur. Though the decadal rate of growth of migrants is lagging behind to the growth of the urban population as both has been

59.3% & 45.2% in decade 1991-2001, 49.5% & 35.8 % in decade 1991-81 respectively. Short Distance migration is considered, people from the other parts of Jaipur district who are coming to Jaipur urban area, migration from other parts of the state is relatively longer distance migration and put in the moderate (medium) distance migration whereas the people from out side the state are in the category of long distance migration. The contribution of the short distance migration in total migration as per census of 2001, it was 17.1% against the 51% were medium distant migrants as they came from other districts of the state and long distance migration from some most contributing states namely Punjab, U. P., & Delhi have there share as 9.6%, 3.3% & 2.3% in total migrants to Jaipur in this same duration. These three states accounted for half of the long distance migration.

These different types of migration spell a meticulous trend over the years. As small distance migration shows a downward trend as its share in total migration which was 28.8% in yr 1981 came to 25.8% in according to census of 1991 and further slipped to 17.1% in census 2001. Medium distance migration exhibited a opposite path to the short distance migration as it advanced to 47% in yr 1991 against 45% in yr 1981 which further ascended to 51% in yr 2001. Contribution of long distance migration in total migration from all states also exhibited rolling down trend. This trend followed suite for the migration from the adjoining states.

COMPOSITIONAL DYNAMICS OF REASON FOR MIGRATION TO JAIPUR URBAN AREA:

Affect of various reasons of migration on peoples of diverse areas is different. Some reasons are more common than others moreover their affect on male and

females is also different. Share of Rural and Urban in-migrants population will widely vary for various cause of migration. Distance of place of origin is also a crucial factor in migration process to any area. Dynamics of various reasons for migration will be analysed from four perspectives.

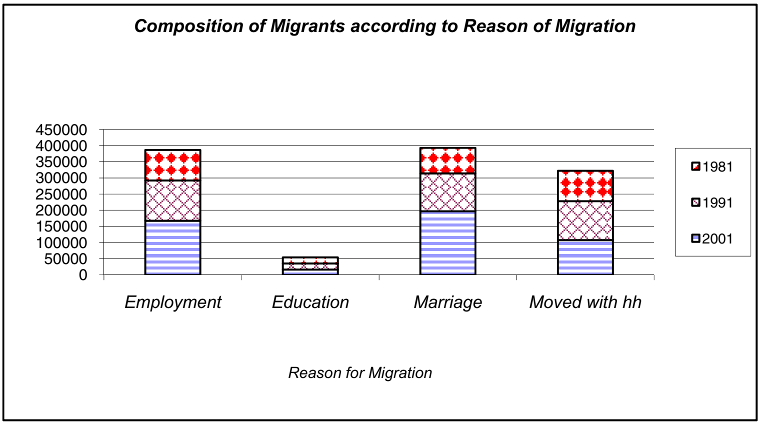
1. Dominance of various reasons for migration;
2. Rural-Urban Paradigm and changes taking place;
3. Gender issues and disparity.

DOMINANCE OF VARIOUS REASONS FOR MIGRATION IN MIGRATION PROCESS:

Person do migrate from a variety of reasons, prominent of them are migration due to 1. Employment 2. Education 3. Marriage 4. Moving with family. Marriage has been the foremost reason for migration as its share in total migrants to Jaipur was 32.1% in yr 2001. People migrating for the employment and/or business with 27.3% contribution in total migration seconded the marriage cause. It was distantly followed by category of persons moving with family with 17.6 % share in total migration. There was a remarkable difference in two dominating categories of people moving due to employment and marriage and it was that the people migrating to Jaipur due to employment is on declining side as it came down to 25% in yr 2001 from 27.3% in yr 1991 and 30.2% in yr 1981 contrary to a gradual increase in people migrating to Jaipur because of marriage as it raised to 32.1% in yr 2001 from 27.8% in yr 1991 and 25.2% in yr 1981.

Education as a cause of migration doesn't have significant contribution in total migration to Jaipur and it is getting meagre over the years. As in yr 1981 its share in

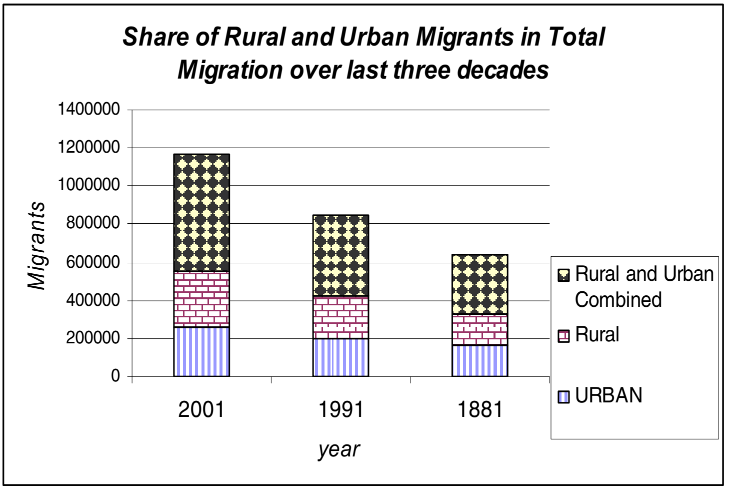
total migration was 6.1% and the figure came to 4.4% in yr 1991 and further dip to 2.7% in yr 2001. This movement is also followed by migrants for education from all the adjoining state, within state and from Jaipur district to Jaipur urban area. People moving with household also followed the decline suite though the rate of decline was steeper than the others as the share of people migrating under this category which was 30.2% in yr 1981 fall to 28.5% in yr 1991 and further it slip to 17.6 % in yr 2001. Composition of various reasons for migration over last three decades is depicted in coming Graph.



RURAL-URBAN PARADIGM:

Intensity of migration widely differs for persons migrating form Rural and Urban areas for various reasons for migration. Flow of migrants from Rural areas to Jaipur outpaced the migrants from urban areas. According to data of census in yr 1981, the share of migrants to Jaipur urban area from rural and urban areas was 53%

& 47 % respectively and this gap remained intact in the coming decades. The trend in rural, urban and combined for last three decades is depicted in graph on next page.



The contribution of rural & urban migrants within a category of reason for migration over last two consecutive decades is tested by calculating the z-values for various category of reason for migration for Rural & Urban areas and significance was tested at 5% level of significance. To test the equality of share of Rural/Urban migrants from any reason of migration over a decade period following hypothesis was set up.

H_0 : Share of Rural (or Urban) migrants due to any reason of migration in a decade is equal. ($p_1=p_2$)

Against

H_1 : $p_1 \neq p_2$

This is tested for two decadal period 1981-9991 & 1991-2001.

where

$$P = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2}, \qquad and\ Q = 1 - P.$$

p_1 is the share of rural/urban migrants due to any reason at a point of time in total migration, p_2 is the share of rural/urban migrants due to that reason after a decade in total migration

To test this hypothesis, Z-value for equality of proportions of migrants from any reason over a decade is calculated and compared with tabulated value at 5% level of significance for the period 1981-9991 & 1991-2001 for rural and urban migrants separately. Four groups according to share of migrants from any reason of migration over a decade period are formed to analyse the Rural-Urban dynamics of the migrant process.

Group1: Share of migrants from any reason of migration from Rural/Urban area over a decade period (in 1981-991 & 1991-2001) is not equal. Means share of peoples migrating from rural & urban areas for a particular reason of migration differ significantly over the period 1981-991 & 1991-2001. Areas falling under this group shows a change in similar direction (i. e. share of urban & rural migrants for that reason of migration has changed considerably over a decade period) for Rural & Urban migrants in terms of their share in total migration for that reason of migration over a decade period.

Group 2: Share of migrants from any reason of migration from Rural/Urban area over a decade period (in 1981-991 & 1991-2001) is equal. Means share of peoples migrating from rural & urban areas for a particular reason of migration don't differ significantly over the period 1981-991 & 1991-2001. Areas falling under this group don't shows any change (i. e. share of urban & rural migrants for that reason of

migration is has not changed over a decade period) for Rural & Urban migrants in terms of their share in total migration for that reason of migration over a decade period.

Group 3: Share of migrants from any reason of migration from Rural area is not equal whereas for migrants from urban areas due to this reason is equal over a decade period (in 1981-991 & 1991-2001). Means share of peoples migrating from Rural areas for a reason of migration differ significantly whereas share of peoples migrating from Urban areas for this reason of migration don't differ significantly over the period 1981-991 & 1991-2001. Areas falling under this group shows different story as share of Urban migrants for any reason of migration in total migration is not equal though for Rural Migrants it is equal over a decade period.

Group 4: Share of migrants from any reason of migration from Urban area is not equal whereas for migrants from Rural areas due to this reason is equal over a decade period (in 1981-991 & 1991-2001). Means share of peoples migrating from Urban areas for a reason of migration differ significantly whereas share of peoples migrating from Rural areas for this reason of migration don't differ significantly over the period 1981-991 & 1991-2001. Areas falling under this group shows different story as share of Rural migrants for any reason of migration in total migration is not equal though for Urban Migrants it is equal over a decade period.

In Group 1 & 2, migration due to any reason from rural and urban areas is in agreement i.e. share of migrants due to any reason over a decade either is significant or insignificant for both rural and urban migrants. In contrary to this In Group 3 & 4, migration due to any reason from rural and urban areas is not in agreement i.e. share

of migrants due to any reason over a decade is significant for urban migrants than it is insignificant for rural migrants or vice-versa.

Z-value for testing hypothesis at 5% level of significance in a group will be as under.

Group 1: Z_u & Z_r>1.96

Group 2: Z_u & Z_r <1.96

Group 3: Z_u >1.96 & Z_r <1.96

Group 4: Z_u <1.96 & Z_r >1.96

Where Z_u and Z_r is the calculated value of Z for migrants due to a reason from Urban & Rural area. The significance of Null hypothesis for all the groups is summarized in table on ensuing page.

Reason for Migration	Contribution of Rural & Urban Migrants over a decade period is in agreement for any reason of Migration			
	Duration 1991-2001		Duration 1981-1991	
	Z_u & Z_r>1.96	Z_u & Z_r <1.96	Z_u & Z_r>1.96	Z_u & Z_r <1.96
Employment	Total Migration, Elsewhere Jaipur District, Gujrat	Haryana, U.P., Delhi	Total Migration, Elsewhere in Jaipur District, in other Districts, Gujarat, Hrayana, U.P., Punjab, Delhi	
Education	Total Migration, Punjab	Gujarat, Haryana, U.P., Delhi	-do-	

Marriage	U.P., Punjab, Haryana, Delhi		Elsewhere in Jaipur District, in other Districts, Gujarat, Hrayana, U.P., Punjab, Delhi	
Moved with Family	Total Migration, Elsewhere in Jaipur District, in other Districts, U.P., Punjab, Delhi		Total Migration, Elsewhere in Jaipur District, in other Districts, Gujarat, Hrayana, U.P., Punjab, Delhi	
Reason for Migration	Contribution of Rural & Urban Migrants over a decade period is not agreement for any reason of Migration			
	Duration 1991-2001		Duration 1981-1991	
	Z_u >1.96 & Z_r <1.96	Z_u <1.96 & Z_r >1.96	Z_u >1.96 & Z_r <1.96	Z_u <1.96 &Z_r >1.96
Employment		in other Districts, Punjab		
Education		Elsewhere in Jaipur District, in other Districts,		
Marriage	Gujarat		Total Migration	
Moved with Family	Haryana			

It is apparent from this summarization that share of rural & urban migrants in the period 1981 & 1991 differ widely for migrants coming from various places.

Especially for migrants coming from other states the share of rural & urban population in yr 1981 & 1991 differ significantly for all the four categories of reason for migration. However this fact was a little bit different in the period of 1991-2001 as migrants coming for education & employment from rural & urban areas of various states don't differ significantly in terms of their share in year 1991 & 2001 in total migration.

Migrants from rural & urban areas due to marriage, employment & education were not in agreement as from some of the areas the proportion of rural migrants in year 1991 & 2001 was significant whereas for urban it was not. Therefore for the duration 1991-2001 migrants from some of the places are not making significant difference in terms of their contribution for some of the reasons to migrate or for rural migrants it is not significant whereas for urban migrants it is significant or vice-versa. This situation was missing in the duration 1981-9991.

GENDER ISSUES AND DISPARITY:

Flow of male and female migration governed by different reasons differently and exhibit a different trait over the years. Looking at total in-migration in Jaipur it is found that contribution of males were phenomenal high in the category of people migrating due to employment and education as against the share of female was higher than males in category of persons migrating due to marriages and moving with family. Moreover the fact of male moving faster than females for employment & education and females move faster than male for marriages and moving along family was also evident in short, medium and long distances migration and this gap at the segregated levels was much explicit than the aggregated level. Following hypothesis was formulated to test the gender disparity in migration.

H_0 : Share of males (or females) migrants due to any reason for migration in a decade is equal (i.e. $p_1=p_2$)

Against

H_1 : $p_1 \neq p_2$

Formula for Z remain same whereas p_1 is the share of male/female migrants due to any reason at a point of time in total migration and p_2 is the share of male/female migrants due to that reason after a decade in total migration.

To test this hypothesis Z-value for equality of proportions of migrants from any reason over a decade is calculated and compared with tabulated value at 5% level of significance for the period 1981-9991 & 1991-2001 for male and female migrants separately. Four groups according to share of migrants of any reason for migrations over a decade period are formed to analyse the Rural-Urban dynamics of the migrant process.

Z-value for testing hypothesis at 5% level of significance for the four groups will be as under.

Group 1: Z_m & Z_f>1.96

Group 2: Z_m & Z_f <1.96

Group 3: Z_m >1.96 & Z_f <1.96

Group 4: Z_m <1.96 & Z_f >1.96

Where Z_m and Z_f is the calculated value of Z for male & female migrants due to a reason. The significance of Null hypothesis for all the groups is summarized in table inserted below.

Reason for Migration	Contribution of Male & Female Migrants over a decade period is in agreement for any reason of Migration			
	Duration 1991-2001		Duration 1981-1991	
	Z_m & Z_f>1.96	Z_m & Z_f <1.96	Z_m & Z_f>1.96	Z_m & Z_f <1.96
Employment	Total Migration, in other Districts	Gujarat, Punjab, Haryana, U.P., Delhi	Total Migration, in other Districts, Gujarat, Punjab,	
Education		Gujarat, Punjab, Haryana, U.P., Delhi	Total Migration	Gujarat, Punjab, Haryana, U.P., Delhi
Marriage	Total Migration,		Total Migration, Elsewhere in Jaipur District, in other Districts, Haryana, Punjab,	
Moved with Family	Total Migration, Elsewhere in Jaipur District		Total Migration, Elsewhere in Jaipur District, in other Districts, Gujarat, U.P, Delhi	
Reason for Migration	Contribution of Male & Female Migrants over a decade period is not in agreement for any reason of Migration			
	Duration 1991-2001		Duration 1981-1991	
	Z_m >1.96 & Z_f <1.96	Z_m <1.96 & Z_f >1.96	Z_m >1.96 & Z_f <1.96	Z_m <1.96 & Z_f >1.96
Employment	Elsewhere in Jaipur District in		Gujarat, Punjab, Haryana, U.P., Delhi	

	other districts			
Education	Total Migration, in other Districts		Elsewhere in Jaipur District, in other Districts	
Marriage		Elsewhere in Jaipur District, in other Districts Gujarat, Punjab, Haryana, U.P., Delhi		Gujarat
Moved with Family		**in other Districts Gujarat, Punjab Haryana, U.P., Delhi**		Punjab, Haryana, U.P., Delhi

It is evident from the above results that the contribution of male and females in different categories over two decades (1981-91 & 1991-2001) has changed considerably and the disparity is widened. As most of the categories in duration (1981-1991) fall in the group where both Z_m & Z_f>1.96 which means proportion of the males & females over a decade was significantly different. In this way male & females for most of the categories were in agreement (Z_m & Z_f >1.96) as both were significant as far as their contribution in total migration over a decade is concerned. Except for the people moving due to education from other states as Z_m & Z_f <1.96 for this category. This means that share male & females migrating due to education from other states in total migration in the year 1981 & 1991 was same and this remained stabilized in year 2001. People migrating due to marriages & moving with family also showed a change in this three decade period as migrating from most of the areas in year 1991 over 1981 exhibited that the share was considerably changed (Z_m & Z_f>1.96) whereas in year 2001 over 1991 it showed that it has not changed for males

though for females it has changed. Thus people moving under these categories have shown a shift in term of increasing share toward females.

Migrants from different areas exhibit a considerable shift in terms of contribution of males or females in total migration over a period of ten years. However overall migrants say that three categories (employment, marriage & moved with family) followed the same suite as the share of male & female was significant for testing hypothesis for equality of the same over the duration 1981-991 & 1991-2001.

Summary:

Contribution of people migrating for education in total migration is on a steep declining as its contribution in total migration has decreased by one third over a two decade period. People migrating due to marriage is showing a phenomenal incremental growth & it is supposed to grow with a faster pace due to decline sex ratio in the city. Migration due to education is having less contribution in total migration and it is going thinner over the years because of education facilities in smaller town and easy accessibility to them in small town. Therefore no longer education is as significant for tempting to migrate as it used to be two decades back. In the coming years this cause of migration will further tend to lose its impact in overall mobility of peoples. People migrating with family is also on a downward trend as people moving with family and due to marriage are together constitute inactive movement as people are not necessarily moving by choice or primarily don't have motive of employment, business or education which itself are related to betterment of life/career.

The share of inactive movements in total migration has came down by 5% over a decade. If this trend continues and the economic progress of the Jaipur

indicates that it will attract the people for economic reasons than the share of migrants in working population will grow which in turns contribute for the economy of the City as the share of people moving with family is declining sharply. Migration from urban areas due to marriages is also getting bigger and voluminous in coming decades this will in turn affect the cultural & social structure of the society and a cosmopolitan culture will emerged.

Analysis of trend of the male & females' migration it can be interpreted that share of employment & education from other states to Jaipur is leading to stabilization & it was not found significant for testing the hypothesis of equality of their share over decades. Whereas for people moving with family the share of males is getting stabilized though for females it was growing. If this scenario continue than growing migration of females in this category will, to some extent, be beneficial to the decreasing sex ratio the city.

Short distance migration which consist the in-migration from various parts of the Jaipur district to Jaipur Urban area is one a sharp decline path in terms of its contribution in total migration. It clearly indicates the tendency of migrating to Jaipur urban area is lower down as periphery of Jaipur urban area is also being developed as its suburb. Better connectivity is raising the number of daily commuter and in near future entire district may be developed as a part of Jaipur urban area and a new Jaipur is shaping up. In such a scenario overall migration to Jaipur urban area from the various parts of Jaipur district will lose its relevance.

References:

Bhagat, R. B. (992). Components of Urban Growth in India with Reference to Haryana: Findings from Recent Censuses, Centre for Training and Research in

Municipal Administration, University of California, Nagarlok, Vol.24, No. 3, 1992, pp. 10-14

Bhattacharya, P. (1988). The informal sector and rural to-urban migration: Some Indian evidence, *Economic and Political Weekly* vol.33(21), 1255-1262.

Jain, M.K., Ghosh, M. and Won Bae, K. *(1993)*. Emerging Trends of Urbanization In India: An Analysis Of 1991 Census Results, *Office of Registrar General, New Delhi.*

Registrar General of India: Census of India (1971, 1981, 1991, 2001) Migration tables Part V A and B, Rajasthan State.

This volume is a collection of five papers/chapters. Two chapters deal with problems in statistical inference, two with inferences in finite population, and one deals with demographic problem. The ideas included here will be useful for researchers doing works in these fields.

ISBN 1-59973-087-1

53995>

9 781599 730875